The Roles of TAL Effectors in Nature in Relation to their Unique Properties as DNA-Targeting Tools

Adam Bogdanove
Cornell University
Ithaca, New York

ajb7@cornell.edu

I will present perspectives on the unique properties of TAL effectors as DNA-targeting tools, and what we can gain from understanding how they work in nature and in plant disease. I'll begin by comparing and contrasting the CRISPR/Cas9 and TAL-effector nuclease (TALEN) systems to highlight some of the unique features of the latter (Figure 1). TAL-effector nucleases work as dimer proteins. Each individual monomer of the TAL-effector domain is about 102 kDa. Targets are encoded by repeats in each protein. In contrast with CRISPRs, typically you need one dimer per target, so it's not readily multiplexed, certainly not as readily as the CRISPR/Cas9. The targeting specificity is typically 15 to 20 bases times 2. Therefore, on the surface, in these base configurations, TAL effectors inherently have great specificity, but, as we heard from Dan Voytas[1] and Ann Ran[2], a number of tricks have been applied within the CRISPR/Cas9 architecture to improve their specificity. So, they are complementary, effective systems.

[1]Pages 29–37.
[2]Pages 69–81.

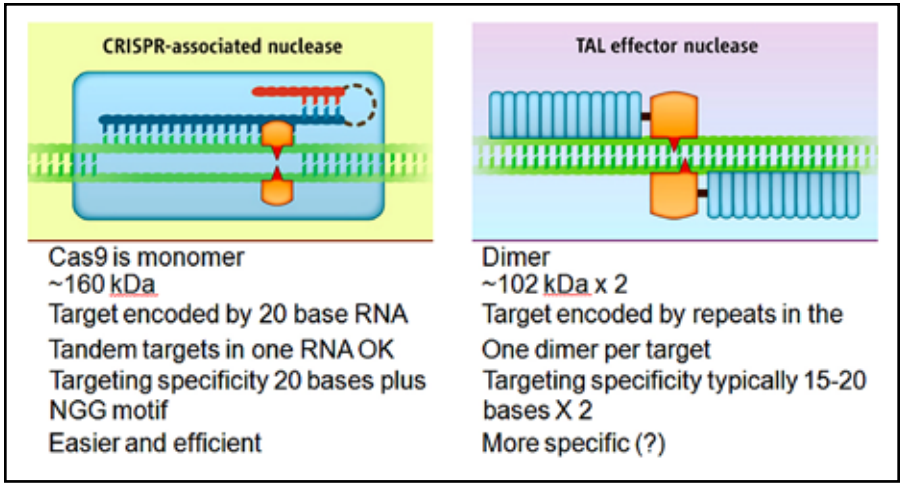

Figure 1. Comparison of DNA-targeting systems.
(Images adopted from van der Oost, 2013)

TALEN Utility

I was interested to see what would happen in the genome-editing field as CRISPRs gained momentum with possibly less emphasis on TAL effectors. However, a proliferation of TALEN kits have been deposited with Addgene[3]—which Dan Voytas[4] and I are familiar with because we deposited there—and over 1,400 requests have been filled for our kit. Also, we and others have web-based tools for TALEN design, which receive thousands of hits per week. Clearly, TALENs remain a popular and, therefore, apparently useful reagent. Figure 2, showing numbers of publications that cite TALENs, reveals a striking rise over the past five years. It remains to be seen whether this trend will continue, but TALENs have been used in all of the organisms of agronomic and agricultural importance in Figure 2. I don't know why anyone would mess with such a thing of beauty and perfection as the catfish, but that has been done. These data exclude progress toward gene-therapy approaches by using TALENs in human-cell cultures.

Certainly, TALENs are a viable and still-popular reagent, but it's interesting to gain more perspective on their unique properties from their native context. We think of them as straight-forward, Lego-block-like modules that can easily be assembled with a one-repeat to one-nucleotide targeting specificity. However, if we step back and think about where they come from, and the selective pressures that are on them, we can gain some interesting insights.

[3]A global, non-profit plasmid repository dedicated to making it easier for scientists to share.
[4]Pages 29–37.

Figure 2. Publications that cite TALENs.

Xanthomonas spp.

The DNA-targeting domain of TALENs comes from the transcription activator-like effectors of *Xanthomonas*, a genus of plant pathogenic bacteria that comprises 20 species within which there are several variants. Collectively, they cause diseases in over 350 plant species. Some of the more economically important diseases are listed in Figure 3. It is noteworthy that not all strains of the pathogenic species deploy TAL effectors and, in those that do, the number of TAL effectors deployed by a pathogenic strain may be anywhere from 1 to 25. There is quite a bit of variability in how consistently bacteria in different contexts actually use TAL effectors as virulence factors to manipulate host-gene expression.

Figure 4 illustrates what TAL effectors do when they function in an important way. They are delivered into the plant cell by the type-III secretion system (T3SS) of the bacterium and, by virtue of some nuclear-localization signals (NLS) in the C-terminal part of the protein, they are imported into the plant-cell nucleus. A translocation signal (T3S) on the N-terminal end gets them out and into the plant cell. A striking feature is an acidic activation domain (AD) on the C-terminal end. That's why they were called transcription activator-like, because they looked to be nuclear localized and they had this sequence at the end that looked like an activation domain typical of a transcription factor. Only in 2007 were they conclusively demonstrated to be translocated transcription factors and to activate gene expression in a very specific way in the host. The targets are recognized by virtue of the repeat region, which directs them to specific locations in the host genome, and then they recruit the transcriptional machinery in some way—yet to be characterized—to drive expression of the downstream gene. That downstream gene in

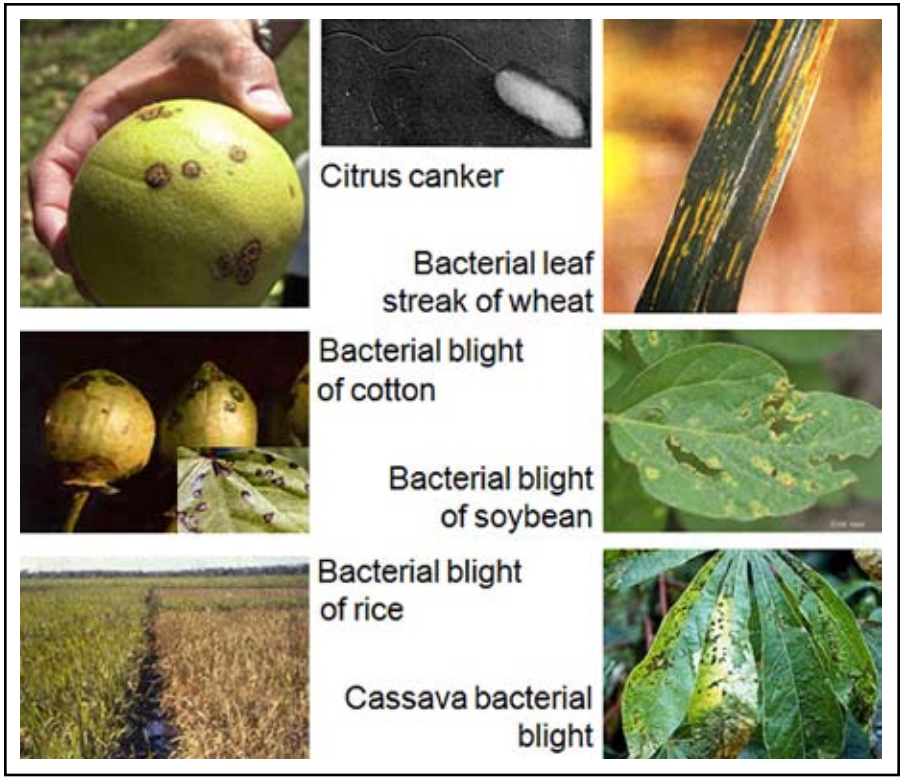

Figure 3. *Xanthomonas* species cause >350 plant diseases.

some way contributes to proliferation of the bacteria for symptom development, and we call it a disease-susceptibility gene, "*S*" in Figure 4. Over evolutionary time, plants have responded in several ways with different types of mechanisms to resist pathogens that are deploying TAL effectors. One that is conceptually quite simple to understand is to acquire polymorphism at the TAL-effector binding site upstream of a major susceptibility gene to prevent activation by the corresponding TAL effector and confer genetically recessive or passive resistance, essentially just taking out the susceptibility (Figure 5).

Another resistance mechanism—discovered in Susan McCouch's lab at Cornell—is polymorphism in a general transcription-factor subunit that provides resistance very broadly to pathogens using TAL effectors (Figure 5), and that's our only genetic evidence for direct interaction of TAL effectors with the transcriptional machinery. The most elegant evolutionary solution to this problem of pathogens with TAL effectors is to juxtapose a TAL-effector binding site upstream of a gene whose activation doesn't confer susceptibility, but triggers a resistant response, in the literature variously called an activation trap or an executor resistance ("*R*") gene—"*Executor*" because it is executing a resistance pathway (Figure 5).

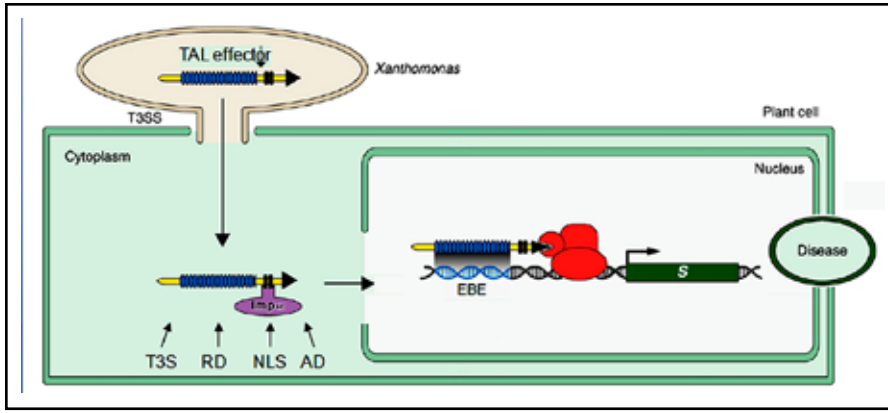

Figure 4. Transcription activator-like effectors are secreted transcription factors that activate host genes to increase susceptibility.
(Bogdanove *et al.*, 2010)

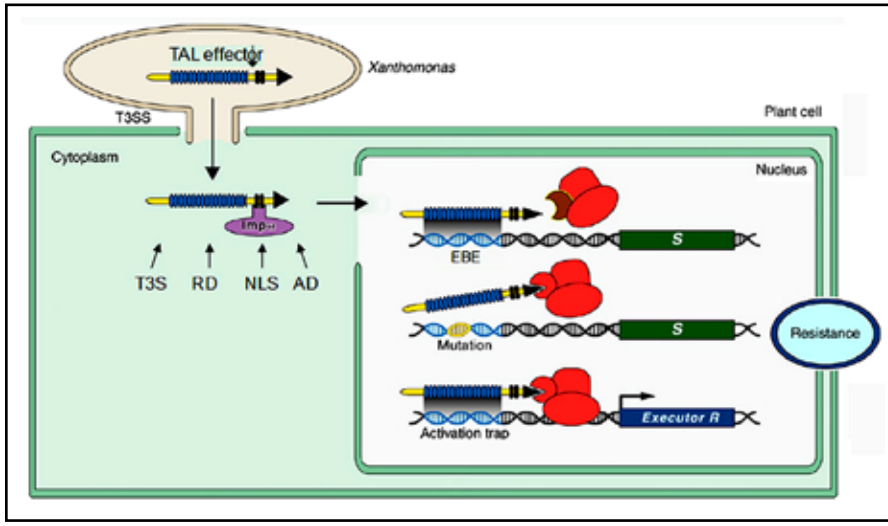

Figure 5. Plants have evolved resistance mechanisms.'

In this context, opposing selective forces act on TAL effectors: (1) for targeting flexibility to accommodate susceptibility-gene polymorphism from one plant variety to the next; and (2) at the same time for stringent specificity to avoid falling into one of these activation traps. If you have evolved a target for an "*S*" gene and you enter a genotype in which there is an "*R*" gene with the same binding site there, or a very similar one, you want to be able to distinguish those. So, in essence, we have two opposing sets of selective forces on TAL effectors, and our working hypothesis is that nature's solution is a modular DNA-recognition mechanism that allows rapid evolution of new specificities by

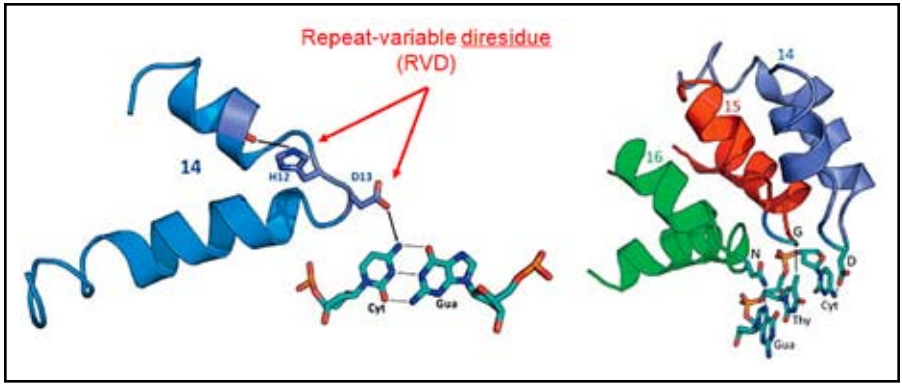

Figure 6. Modules are 33–35 amino-acid repeats. Each forms a 2-helix bundle and recognizes a third base. (Mak *et al.*, 2012)

recombination of those modules, but also—as discussed below—allows a tunable targeting specificity, not only specificity in itself in a qualitative fashion, but the stringency of that specificity can be modulated in a quantitative fashion.

TAL-Effector Modules

The TAL-effector modules are 33 to 35 amino acid repeats and any given TAL effector may have between 5 and 30 of these repeats in the central repeat region (Figure 6). Each of these repeats forms a two-helix bundle and each repeat recognizes a single nucleotide, so, continuously, the number of repeats and the composition of those repeats define the number of nucleotides and the composition of that nucleotide stream in the target. The repeat variable diresidue (RVD)—two amino acids—resides on an inter-helical loop and the 13th residue side chain reaches out and makes base-specific contact. Residue 12 reaches back and, through interactions with the first helix, shapes the conformation of the loop to affect DNA-binding specificity.

If you stack these repeats up in a lateral-like fashion to track the DNA, each repeat interacts with one base and the repeats all together assemble into a superhelix that wraps the DNA (Figure 7) for interactions in the major group with each of those RVDs and their corresponding nucleotides.

RVD-Nucleotide Interaction

Figure 8 illustrates the four major repeat types with each of the most common RVDs that are used in genome editing. A straightforward correspondence exists between the RVD sequences and the nucleotides that they prefer, but even though people typically use NI for A and HD for C and NN for G or A and NG for T, the heights of the letters in Figure 8 tell us how often each RVD is found associated with that nucleotide in nature. The stringency is not entirely strict, so that NI is sometimes found in association with C and can also be happy in an array opposite a C. So, there is some subtlety already in the major RVD-nucleotide interactions. However, beyond that, there is a range of RVDs that actually occur in nature that weren't characterized until recently. A group at Peking

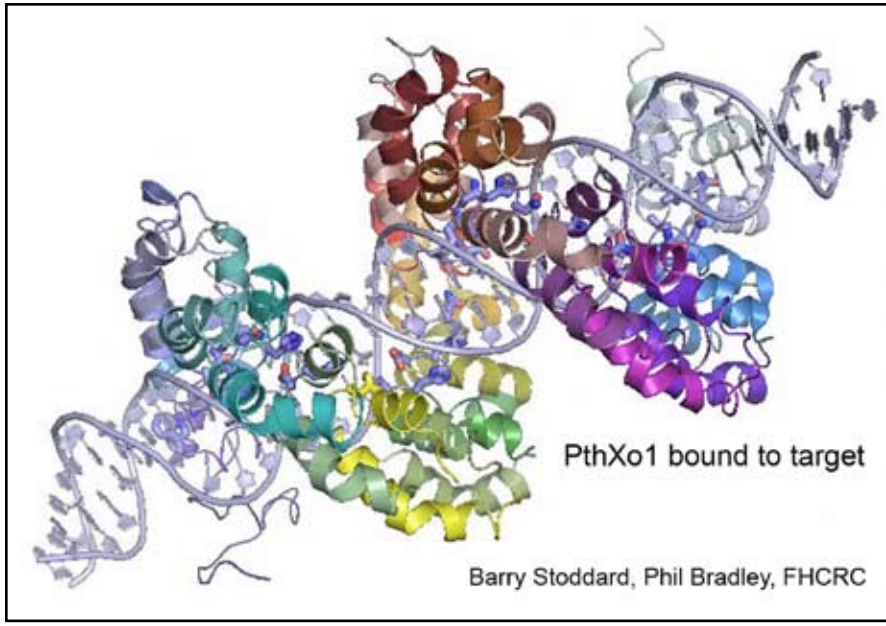

Figure 7. Repeats assemble into a superhelix that wraps the DNA.
(Mak *et al.*, 2012)

University tested all possible RVDs—all amino acid residues in the 12th position and all the amino acid residues in the 13th position—assayed by using a TAL effector as a transcriptional activator, in a human cell line where they varied three repeats as a homomeric triplet (Figure 9). They tested each of these different RVDs as a homomeric triplet in the middle of the repeat array on substrates with corresponding triplets of either A or T or C or G. Each of the quadrants in the array shows activity for A or T or C or G according to the colors, green, red, blue or yellow, respectively, and the gradient schemes (0 to ≥20).

For example, at HD the specificity is recapitulated with very strong activity on C and virtually no activity on the other bases (Figure 9). At NN, the one for dual specificity, it's recapitulated very nicely in their assay as well, you see G and A. And NK—which was discovered early as a substitute for NN that gives specificity for G—is very specific for G, but the activity is low relative the NN variant. And then there are weird ones, like RV, which has not actually been observed in nature, but might be useful as a wildcard RVD; it shows good activity for any of the four bases. So, this pool of unexploited RVD variation exists in TAL effectors found in nature.

Frameshift Accommodation

An elegant study was done by colleagues in Germany and France who were puzzled by the presence of the atypical repeats in TAL effectors. A typical repeat is 34 amino acids, with the RVD at 12 and 13 (Figure 10). But occasionally you see these ones with an insertion at the C terminal part of the repeat, or a deletion of the very C terminal end of the repeat or an insertion in the N-terminal part of the repeat. They wondered if these

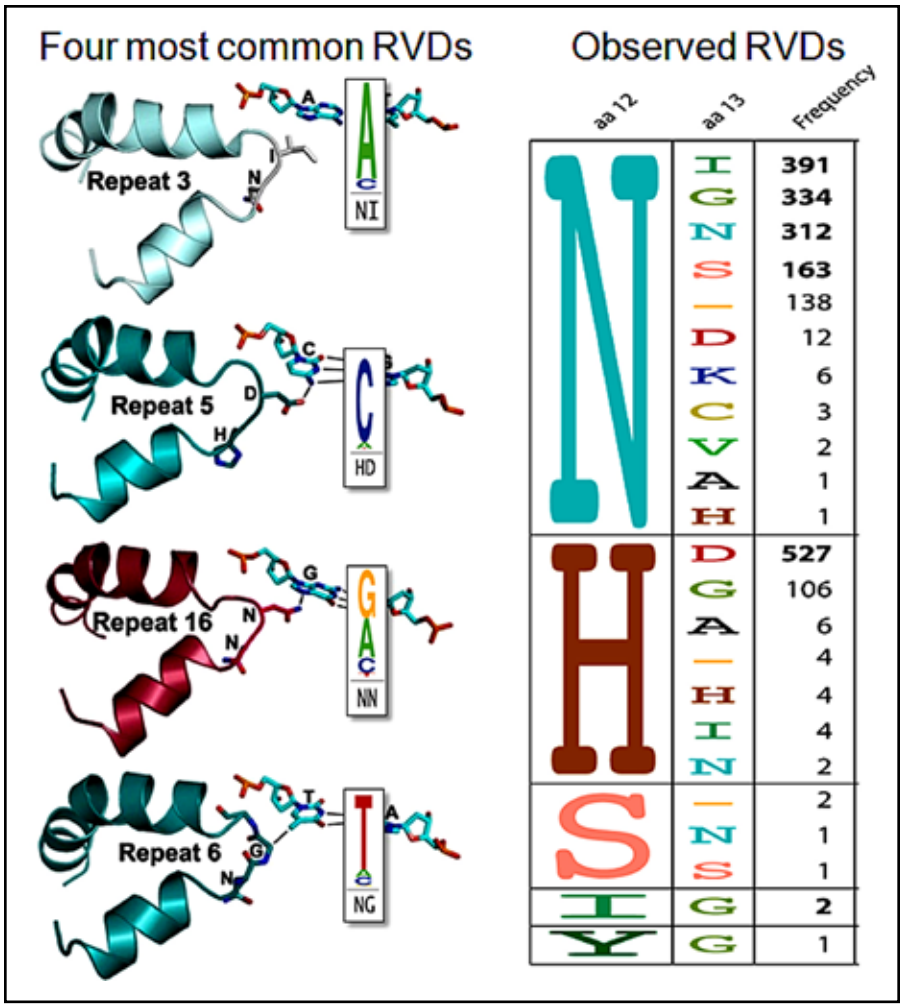

Figure 8. Major repeat types and common RVDs.
(Moscou and Bogdanove, 2009; Boch and Bonas, 2010)

are doing anything different. It turns out that they provide a frameshift accommodation. If you take an array of standard 34 amino acid repeats, and you put it on a template, then if you pop out a purine, for example, you move the template and TALEN-effector register out of frame (Figure 10), so binding dissolves after this point and you lose activity entirely. They found that the presence of the aberrant repeats allows a disengagement of the repeat from the interaction and accommodation of that frameshift. So all those downstream shift back up and then match the template (Figure 10). So far, this is completely unexploited in DNA-targeting applications, but it's a very interesting feature of TALEN effectors occurring in nature.

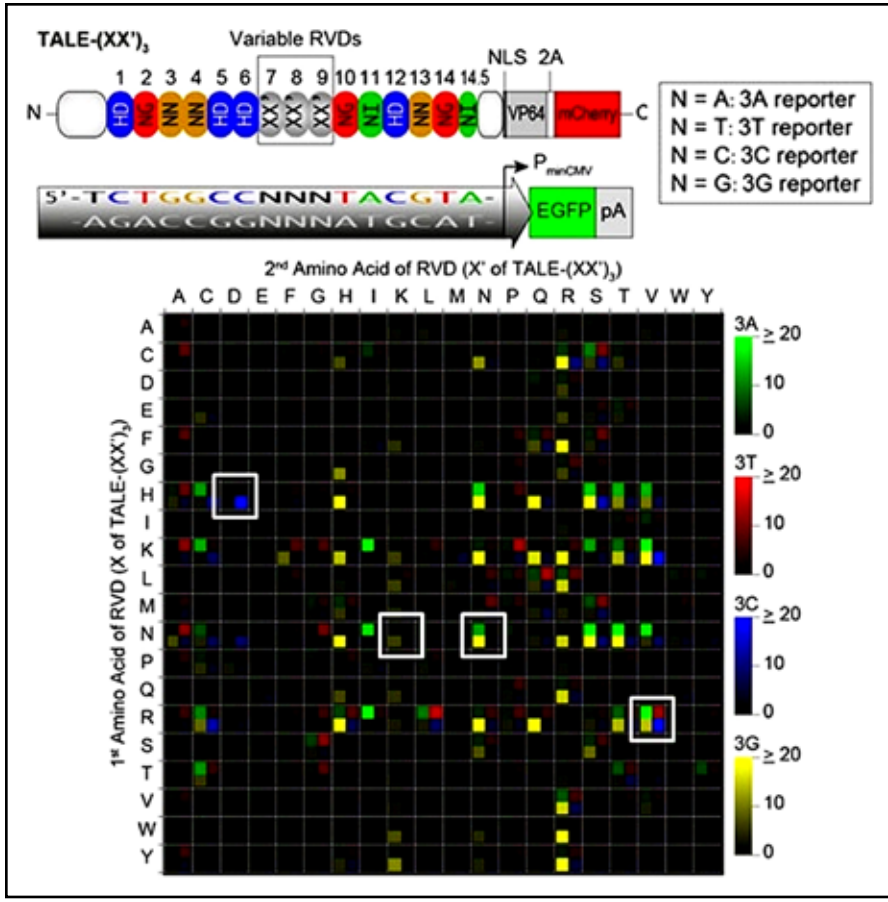

Figure 9. All possible RVDs.
(Yang *et al.*, 2013)

Array Length

Another curiosity that we've explored recently is the effect of variation in the length of the array, the number of the repeats in the array, on specificity and activity. If you look simply at a collection of TAL effectors in nature, there is quite a variation in the number of repeats, as mentioned before. Some are probably pseudogenes, essentially just 2 RVDs long, and some are as long as 34 RVDs but the peak—the most common in nature—is somewhere between 16 and 20 RVDs (Figure 11). This is also the most common length used in most DNA-targeting applications. To nail down whether this finding is functionally significant, Fabio Rinaldi, a postdoc in my lab, assisted by undergraduate Ava Fan, took on the question of the relationship of length to affinity and specificity. Figure 12 shows the experimental set up. They took, as an anchor, the first 10 RVDs of the

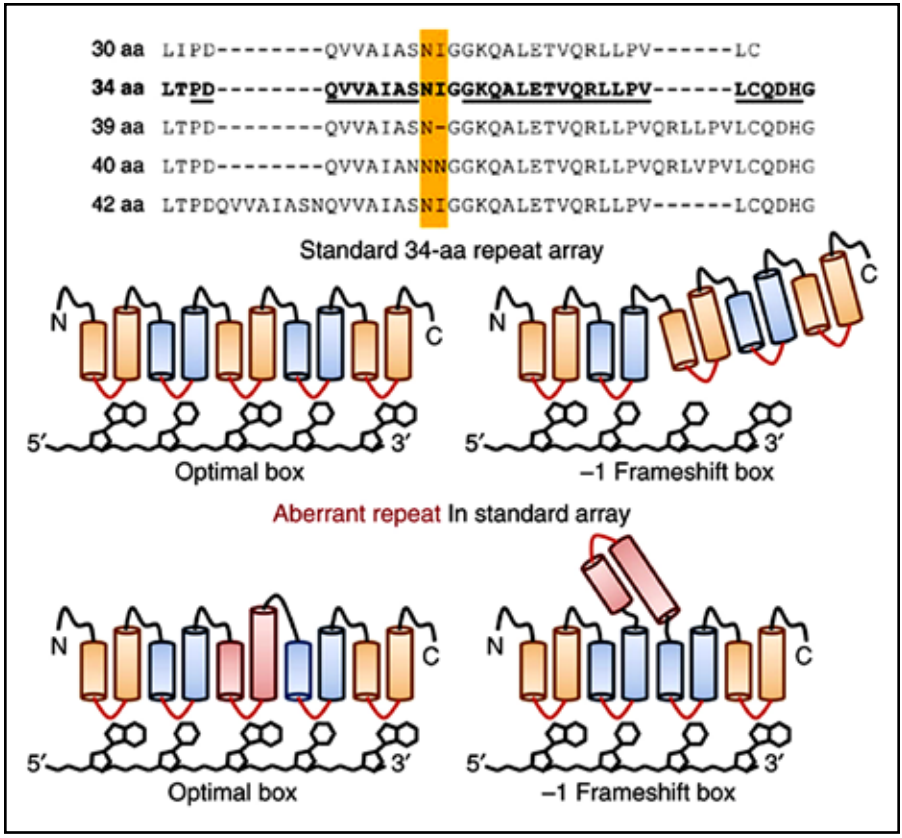

Figure 10. Atypical repeats accommodate single-bp deletions.
(Richter *et al.*, 2013)

well studied TAL effector PthXo1 and then they used repeating quadruplets of the four major RVDs. So, they ended up with a 10-mer, a 14-mer, an 18-mer, a 22-mer and a 26-mer, which they tested on substrates that matched perfectly and had corresponding repeating subjects of A, C, G, and T. So each was anchored by the same register and the effects of lengthening the array on specificity and affinity could be assessed. As a control, they tested interactions of these proteins also on a scrambled-DNA target representing nonspecific DNA interaction.

A lower dissociation constant means stronger binding and higher affinity. It turns out that when you measure the 10-, 14-, 18-, 22- and 26-mer arrays on the target DNA, a drop in the Kd occurs, with a diminishing-return scenario after about 18 RVDs (Figure 13); the gain in overall affinity for lengthening the array is diminished. In contrast, the non-specific interaction on the scrambled DNA showed that affinity drops in a more linear fashion. The ratio of the affinity on the non-specific DNA to the affinity on the specific DNA produced a gaussian-like distribution or bell curve (Figure 14). And, interestingly

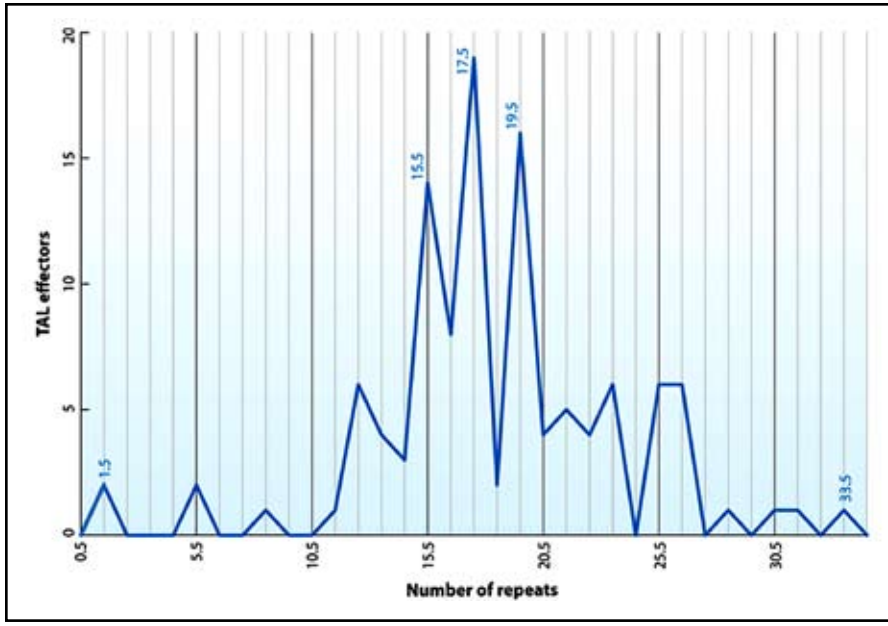

Figure 11. Frequency of array length in nature.
(Boch and Bonus, 2010)

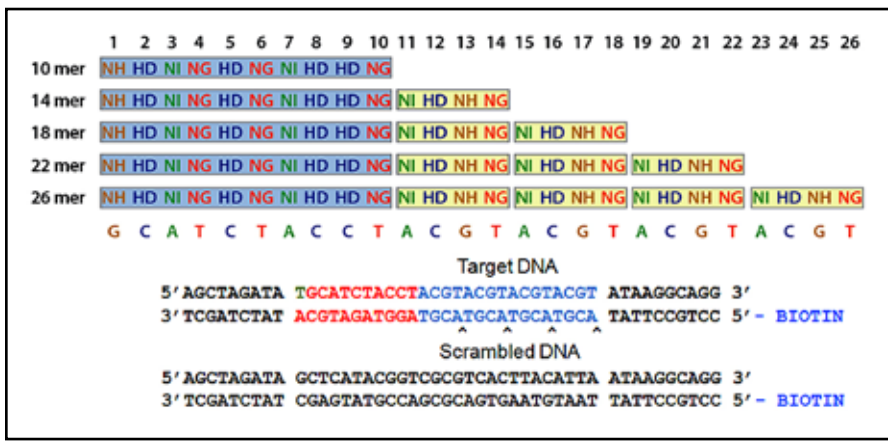

Figure 12. Relationship of length (# repeats) to affinity and specificity.

then, this ratio represents the overall specificity. And we see there is a clear optimum somewhere between 16 and 20. In nature, it seems that the most commonly observed length of TAL effectors is the one that gives the best balance between targeting affinity and minimal off-targeting.

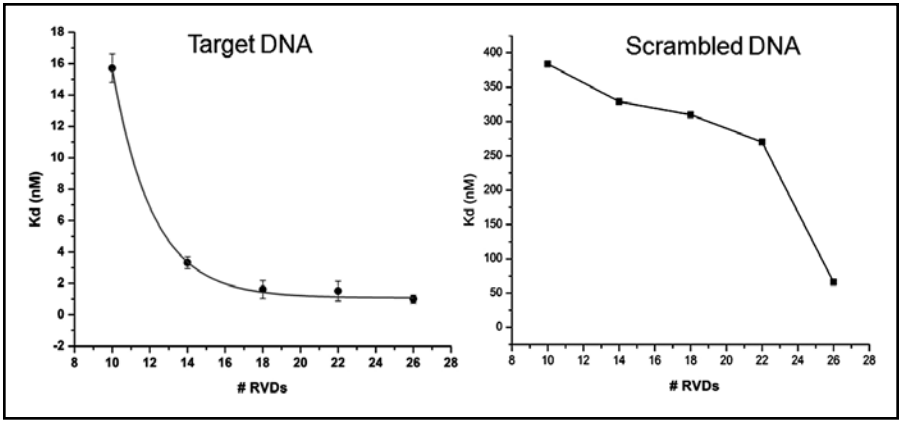

Figure 13. Size versus affinity.

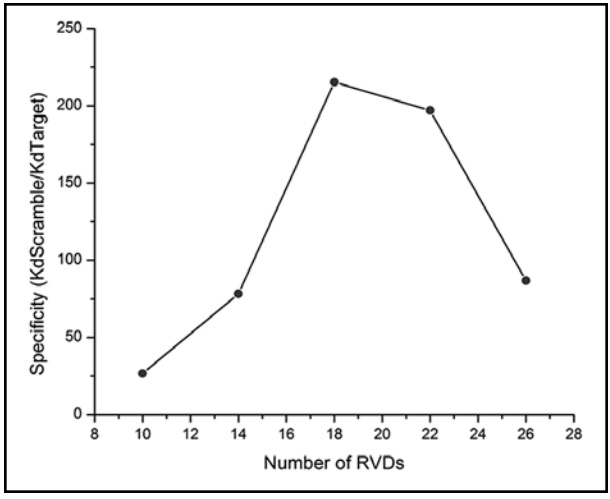

Figure 14. Size versus specificity (Kd_{scr}/Kd_{tar}).

To summarize those aspects, it's fair to say that with varying RVD composition and varying numbers of repeats and depending on the presence of these aberrant repeats that I mentioned, affinity and specificity of a TAL effector array can vary. And when I'm talking about specificity, it's important to keep in mind not just qualitative specificity but also quantitative.

From Nature

I want to wrap up with an observation from nature that Bing Yang[5] already touched on. We haven't exploited this technology to the extent that we can in DNA-targeting applications,

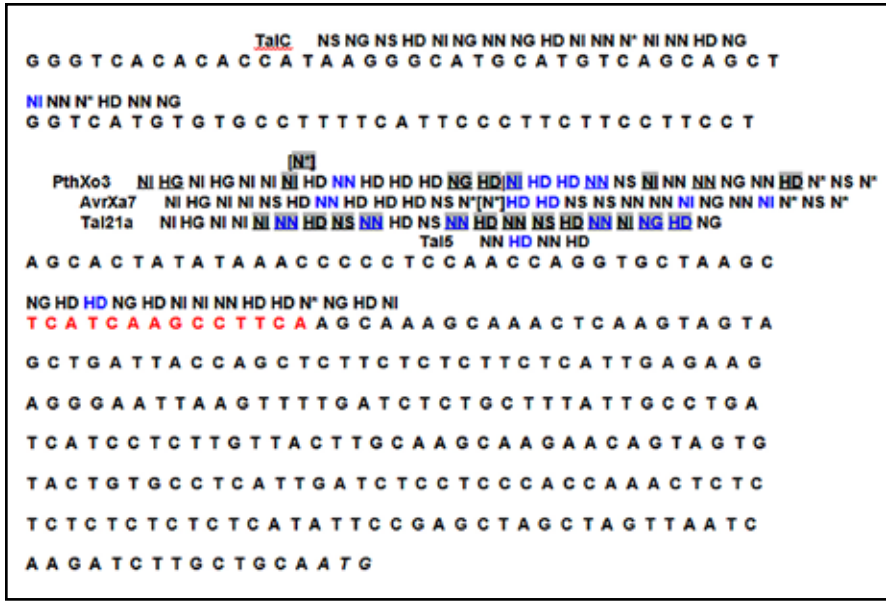

Figure 15. Distinct TAL effectors from different strains target the rice bacterial blight *S* gene *SWEET14*.
(contributions from many research groups)

but to what extent does evolution exploit these adaptations? In the context of bacterial blight of rice disease that Bing introduced, if we look at one of the susceptibility genes, *SWEET14*, it turns out that there are at least five distinct TAL effectors from different strains that go after this gene promoter. And the variation across these is very telling. We have some that are fairly long, some that are fairly short, some that target the same general sequence and others that target distinct sequences. And in the group to focus on, I've underlined differences in proteins PthXo3 and the Tal21A, from protein AvrXa7 (Figure 15). The blue colors show where there are mismatches, alignments of RVDs opposite their non-preferred nucleotides. Something special about AvrXa7 is that it is caught by an activation trap in another genotype. It triggers resistance in rice that carries the gene Xa7. Tal21A and PthXo3 don't fall into that trap; they target the same spot in the susceptibility gene, but they avoid the activation trap. The differences here are not so great between AvrXa7 and Tal21A; it's essentially a deletion of four repeats at the end and then several modifications, which maintain activity on this susceptibility-gene substrate, but clearly interfere with binding to the activation trap. But with PthXo3, additions and a couple of differences also help this protein to avoid the activation trap. The little bracketed RVD is one of those aberrant types, so it can accommodate the sequence by looping out this particular repeat and the presence of that repeat combined with those other differences

[5]Pages 53–59.

clearly is exploited to prevent getting trapped by Xa7. It's fair to say just from this simple example that adaptations are exploiting these unique properties of TALEN effectors.

My take-home message is that we are continuing to better understand the subtle properties of TAL effectors in part by studying them in their native context, but we are very far from having fully exploited them in DNA-targeting applications.

References

Boch J Bonas U (2010) Xanthomonas AvrBs3 family-type III effectors: Discovery and Function. Annual Review of Phytopathology 48 419–436.

Bogdanove A *et al*. (2010) TAL effectors: Finding plant genes for disease and defense. Current Opinion in Plant Biology 13 394–401.

Mak AN-S *et al*. (2012) The crystal structure of TAL effector PthXo1 bound to its DNA target. Science **335** 716–719.

Moscou MJ Bogdanove AJ (2009) A simple cipher governs DNA recognition by TAL effectors. Science 326(5959) 1501.

Richter A *et al*. (2014) A TAL effector repeat architecture for frameshift binding. Nature Communications 5 3447. doi:10.1038/ncomms4447.

van der Oost J (2013) Molecular biology. New tool for genome surgery. Science 339(6121) 768–770.

Yang J *et al*. (2014) Complete decoding of TAL effectors for DNA recognition. Cell Research 24 628–631.

ADAM BOGDANOVE is a professor of plant pathology and plant-microbe biology at Cornell University. His research centers on diseases of rice caused by the bacterium *Xanthomonas oryzae*, with a special focus on TAL effectors, *i.e.* transcription factors injected by the bacterium into the plant to activate host genes during infection. He discovered the modular mechanism by which TAL effectors recognize their DNA targets and has pioneered the use of TAL effectors as customizable DNA-targeting domains for a variety of applications including genome editing and synthetic biology.

DR. BOGDANOVE earned his BS degree in biology at Yale University in 1987, and his PhD in plant pathology at Cornell in 1997. Following postdoctoral work at Purdue University and later at the Boyce Thompson Institute, he joined the Department of Plant Pathology at Iowa State University in 2000 and moved to Cornell in 2012.